

Weekly Report

20171023-20171030

Han Dongming Oct. 30 2017

处理数据

用美国那边提供的算法去找网络中异常的数据结构。但是整个过程超级慢，数据才几千个点，就要跑一天，很难做到实时查询啊，而且算法基于matlab改起来也需要经验，后端又是python，很难接入，只能预先跑出，再找可能为异常的结构的时候，还是需要用人眼去看实际的结构样子，更多的是看力引导之后的拓扑结构。

综述

准备写关于Anomaly detection的综述

关于投稿

周六组会朱博报告的那一篇的套路，的确可以借鉴一下，通过一个其他领域的算法去搜小搜索空间，然后再可视化出来，view视图也可以套路一下，例如报告那一篇：

1. 高维到二维的投影(聚类)
2. 对比
3. 原始数据
4. 细节总览

paper

Understanding a sequence of sequences:

Visual exploration of categorical states in lake sediment cores

从时序数据中分析相似的pattern,三种方法

1. 最长公共子序列
2. 最少的修改cost可以让两个一样

3. 子集构建占对方的比例

$$J(\Omega_1, \Omega_2) = \frac{|\Omega_1 \cap \Omega_2|}{|\Omega_1 \cup \Omega_2|} = \frac{|\Omega_1 \cap \Omega_2|}{|\Omega_1| + |\Omega_2| - |\Omega_1 \cap \Omega_2|}$$